GENE AND PROTEIN SEQUENCES OF PHAGE T4 *gene 35*

1. INTRODUCTION

The present invention relates to nucleotide sequences of bacteriophage T4 *gene 35*

5   and amino acid sequences of its encoded protein, as well as derivatives and analogs thereof
and antibodies thereto. The present invention further relates to the use of nucleic acids
encoding bacteriophage T4 *gene 35* and its encoded protein, as well as derivatives, and
analogs thereof, in the construction of nanostructures, *i.e.*, nanometer sized structures useful
in the construction of microscopic and macroscopic structures.

10

2. BACKGROUND OF THE INVENTION

Bacteriophage, viruses that attack bacteria, are generally composed of a protein coat
which surrounds genetic material. Bacteriophage T4, a T-even phage, consists of an
icosahedron shaped head which contains DNA, a tail (a hollow cylinder of contractile

15   protein) which serves as an injection tube of the DNA and tail fiber appendages which
emanate from the base of the tail. The tail fibers serve to attach the phage to the bacterial
surface in a process known as adsorption.

Bacteriophage T4 tail fiber is composed of four non-covalently joined parts in a
stiff, heat stable, protease resistant structure. This structure can be represented

20   schematically as follows (N= amino terminus, C= carboxy terminus): N[gp34
homooligomer]C - N[gp35]C - N[gp36 homooligomer]C - N[gp37 homooligomer]C. The
gp34 homooligomer ("P34"), gp36 homooligomer ("P36"), and gp37 homooligomer
("P37") are rod-shaped structures in which two identical β sheets, oriented in the same
direction, are fused face-to-face by hydrophobic interactions between the sheets juxtaposed

25   with a 180° rotational axis of symmetry through the long axis of the rod. gp35, by contrast,
is a monomeric polypeptide that attaches specifically first to the N-terminal region of the
P36 homooligomer and then to the C-terminus of the P34 homooligomer and forms a joint
between these two rods having an average angle of 137° (±7°) or 156° (±12°).

The self assembly of the tail fiber is regulated by a predetermined order based on the

30   interaction of specific protein subunits whereby structural maturation caused by formation
of the first subassembly permits interaction with new (previously disallowed) subunits.
During T4 infection of *E. coli*, gp37 (the monomeric 109 Kda translation product of *gene
37*) forms the homooligomer P37, with the aid of 2 accessory (chaperon) proteins, gp57 and

gp38; this process is believed to initiate near the C-terminus of gp37. Once P37 is formed, the N-terminus of P37 initiates the oligomerization of two gp36 molecules of 23 Kda each, in a butt-end joint to form the P36 homooligomer rod. The N-terminus of P36 then attaches to the carboxy terminal region of a gp35 monomer; this interaction stabilizes P36 and forms

5   the flexible angle joint of the tail fiber. The amino terminal region of gp35 then attaches to the C-terminus of P34 (the homooligomerization of which requires the chaperon protein gp57). This regulation of self assembly of the tail fiber by a predetermined, ordered interaction of specific subunits results in the production of a structure of exact specifications from a random mixture of the tail fiber subunit components.

10 While the strength of most metallic and ceramic based materials derives from the theoretical bonding strengths between their component molecules and crystallite surfaces, it is significantly limited by flaws in their crystal or glass-like structures. These flaws are usually inherent in the raw materials themselves or developed during fabrication and are often expanded due to exposure to environmental stresses.

15 The emerging field of nanotechnology has made the limitations of traditional materials more critical. The ability to design and produce very small structures (*i.e.*, of nanometer dimensions) that can serve complex functions depends upon the use of appropriate materials that can be manipulated in predictable and reproducible ways, and that have the properties required for each novel application.

20 Biological systems serve as a paradigm for sophisticated nanostructures. Living cells fabricate proteins and combine them into structures, such as bacteriophage tail fibers, that are perfectly formed and can resist damage in their normal environment. In some cases, such as with bacteriophage tail fibers, these structures are created by a process of self-assembly, the instructions for which are built into the component polypeptides. These

25 natural proteins are also subject to proofreading processes that insure a high degree of quality control. Advantages of using natural proteins to construct nanostructures are that the resulting structures are stiff, strong, stable in aqueous media, heat resistant, protease resistant, and can be rendered biodegradable. Additionally, large quantities of nanostructure parts and subassemblies can be easily fabricated in microorganisms and stored and used as

30 needed.

There is a need in the art for methods and compositions that exploit these unique features of proteins to form constituents of synthetic nanostructures. The need is to design

materials that have properties which can be tailored to suit the particular requirements of nanometer-scale technology. Moreover, since the subunits of most macrostructural materials, ceramics, metals, fibers, etc., are based on the bonding of nanostructural subunits, the fabrication of appropriate subunits without flaws and of exact dimensions and

5 uniformity should improve the strength and consistency of these macrostructures because the surfaces are more regular and can interact more closely over an extended area than larger, more heterogeneous material.

The use of bacteriophage tail fiber components in the construction of nanostructures is further described in PCT Publication WO 96/11947, dated April 25, 1996, the contents of

10 which are incorporated herein in its entirety.

Phage T4 *gp35* is located between genes *gp34* and *gp36*. A sequence for *gp35* is available on the NCBI database (NCBI.NIH.GOV) within the sequence T4g34-t (bases 4188-5075). The T4g34-t sequence reveals that *gene 35* has an open reading frame, ORF35, that is predicted to encode a protein having a molecular weight of 32,334 Daltons.

15 The NCBI database also predicts an open reading frame, ORF34.1, that extends 241 nucleotides between genes *gp34* and *gp35*, and encodes a deduced protein having a molecular weight of 7,334 Daltons (in a different reading frame from ORF35).

The discrepancy between the gp35 molecular weight of 32,334 Daltons predicted by the NCBI sequence and that of 39,000-40,000 Daltons reported from SDS-polyacrylamide

20 gel electrophoresis (SDS-PAGE) analysis, has previously been acknowledged (Karam, J. (ed.), 1994, Molecular Biology of Bacteriophage T4, ASM Press, Wash. D.C., pp. 491-514 at 514).

Citation of a reference hereinabove shall not be construed as an admission that such reference is prior art to the present invention.

25

3. SUMMARY OF THE INVENTION

The present invention relates to nucleotide sequences of bacteriophage T4 *gene 35*, and amino acid sequences of the encoded bacteriophage T4 gene 35 protein, as well as derivatives (*e.g.*, fragments) and analogs thereof, and antibodies thereto. The present

30 invention further relates to nucleic acids hybridizable to or complementary to the foregoing nucleotide sequences, as well as equivalent nucleic acid sequences encoding a bacteriophage T4 gene 35 protein.

- 3 -

The present invention also relates to expression vectors encoding a bacteriophage T4 gene 35 protein, derivatives or analogs thereof, as well as host cells containing the expression vectors encoding the bacteriophage T4 gene 35 protein, derivative or analog thereof. As used herein, "*gene 35 (gp35)*" shall be used with reference to the bacteriophage

5   T4 *gene 35*, whereas "gene 35 (gp35)" shall be used with reference to the protein product of bacteriophage T4 *gene 35*.

The present invention also relates to methods of production of the gp35 proteins, derivatives and analogs, such as, for example, by recombinant means.

The invention further relates to gp35 proteins, derivatives (*e.g.*, fragments), and

10   analogs having an angle joint domain that has been modified so as to form average angles different from the natural average angle of 137° (±7°) or 156° (±12°).

The invention also relates to gp35 proteins, derivatives and analogs which exhibit thermolabile interactions with tail fiber binding partners.

The invention further relates to gp35 derivatives and analogs which are functionally

15   active, *i.e.*, they are capable of displaying one or more known functional activities associated with a full-length (wild-type) gp35 protein. Such functional activities include, but are not limited to, antigenicity [ability to bind (or compete with gp35 for binding) to an anti-gp35 antibody], immunogenicity (ability to generate antibody which binds to gp35), and ability to bind (or compete with gp35 for binding) to a ligand for gp35, and ability to

20   multimerize with other phage products such as P34 and/or P36.

The gp35 protein, derivative or analogs thereof disclosed herein may be used for the production of anti-gp35 antibodies which antibodies may be used diagnostically in immunoassays for the detection or measurement of gp35 protein.

The invention also relates to fragments (and derivatives and analogs thereof) of gp35

25   which comprise one or more domains of a gp35 protein, *e.g.*, the P34 or P36 binding domain, and/or retain the antigenicity of a gp35 protein (*i.e.*, are able to be bound by an anti-gp35 antibody).

The present invention further relates to the use of nucleotide sequences of *gp35* and its encoded amino acid sequence in the construction of nanostructures, *i.e.*, nanometer sized

30   structures useful in the construction of microscopic and macroscopic structures.

4. DESCRIPTION OF THE FIGURES

- 4 -

Figures 1A-1B. T4 bacteriophage. Schematic representation of the T4 bacteriophage particle (Figure 1A), and a schematic representation of the bacteriophage T4 tail fiber (Figure 1B).

Figure 2. Sequence of bacteriophage T4 gp35. The nucleotide (SEQ ID NO:1) and
5 deduced amino acid (SEQ ID NO:2) sequences of bacteriophage T4 gp35. The gp35 protein sequence shown in Figure 3 (encoded by nucleotides 4,127-5,011 of Figure 3) lacks amino acid numbers 1-77 of Figure 2. Amino acid numbers 1-7, 18-56 and 65 of Figure 2 appear as part of the ORF34.1 sequence in Figure 3 (encoded by nucleotides 3,894-4,088 of Figure 3).

10 Figure 3. NCBI database sequence containing bacteriophage T4 *gene 34*, *gene 35* (with errors), *gene 36* and *gene 37*. The nucleotide sequence containing *gene 34*, *gene 35* and *gene 36* (SEQ ID NO:3) and the amino acids encoding the gene products of gene 34 (SEQ ID NO:4; ORF 34.1, SEQ ID NO:5) gene 35 (SEQ ID NO:6), GENE 36 (SEQ ID NO:7) and gene 37 (SEQ ID NO:8).

15

5. DETAILED DESCRIPTION OF THE INVENTION

The present inventor has discovered that significant errors are present in the nucleotide and amino acid sequences of gp35 disclosed in the prior art. Indeed, the inventor has discovered that the prior art predicted amino acid sequence of gp35 lacks 77 amino acid
20 residues at the N-terminus of the actual protein and that 15 of the 16 amino acid residues corresponding to the N-terminal residues of the prior art predicted gp35 are incorrect. The invention thus provides sequences of gp35 that correct these prior art errors.

The present invention thus relates to nucleotide sequences of *gp35* and amino acid sequences of encoded gp35 proteins, as well as derivatives and analogs thereof, and
25 antibodies thereto.

As described by way of example *infra*, the present inventor has isolated and characterized the gene encoding bacteriophage T4 gp35, a tail component necessary for the formation of bacteriophage T4 tail fibers. The nucleotide sequence encoding gp35 was determined to be distinct from that previously reported in the NCBI database (Figure 3).
30 According to the present invention, the *gp35* nucleotide sequence encodes a protein that has a different N-terminus and a molecular weight that is 24% greater than that predicted by the sequence in the NCBI database (nucleotides 4,127-5,011 of Figure 3).

- 5 -

In contrast to the prior art, by providing the correct sequence of gp35 (including the correct amino-terminal portion of the molecule), the present invention enables recombinant production and genetic manipulation of the gp35 protein.

In a preferred aspect, the present invention provides a purified bacteriophage gp35

5   protein that is not contained in a gel (e.g., a gel suitable in which to conduct electrophoresis).

In a specific embodiment, the invention relates to a composition comprising at least 1, 10, 50, 100 or 500 nanogram(s), 1, 10, 50, 100 or 500 microgram(s), or 1, 10, 50, 100 or 500 milligram(s), of purified non-denatured gp35 protein.

10   The *gp35* gene sequence of the invention can be a naturally occurring sequence or in variant form, whether natural, synthetic, or recombinant. In a specific embodiment, the gp35 protein is not native (i.e., not naturally occurring).

In a specific embodiment, the present invention relates to a bacteriophage T4 gp35 protein variant containing the amino acid sequence depicted in Figure 2 (SEQ ID NO:2)

15   wherein only conservative substitutions relative to the sequence in Figure 2 are made. The invention also relates to purified molecules comprising bacteriophage T4 gp35 protein fragments, which fragments consist of at least the amino acid sequence depicted in Figure 2 (SEQ ID NO:2) from amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79 or 81-93, as well as derivatives thereof, e.g., in which only conservative substitutions

20   relative to the sequence in Figure 2 are made. Nucleic acids encoding such proteins, and their complement, are also within the scope of the invention.

The invention additionally relates to proteins, derivatives, fragments or analogs containing an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85% or 90% identity to amino acids number 1 to 100 in Figure 2 over a 100 amino acid sequence. As

25   used herein, amino acid sequence homology refers to amino acid sequences having identical amino acid residues or amino acid sequences containing conservative changes in amino acid residues. In another embodiment, a gp35 homologous protein is one that shares the foregoing percentages of sequences identical with the naturally occurring gp35 protein over a 100 amino acid length.

30   The invention additionally relates to proteins, derivatives, fragments or analogs containing an amino acid sequence having at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85% or 90% identity to amino acids number 57 to 93 in Figure 2

over a 36 amino acid sequence. In another embodiment, a gp35 homologous protein is one that shares the foregoing percentages of sequences identical with the naturally occurring gp35 protein over a 36 amino acid length.

The invention also relates to proteins encoded by nucleic acids hybridizable to a

5 *gp35* gene under non-stringent, moderately stringent, or stringent conditions. In a specific embodiment, such a protein is encoded by a nucleic acid hybridizable to a DNA having a nucleotide sequence consisting of the coding region of SEQ ID NO:1 or its complement.

As defined herein, a gp35 derivative may be a fragment or amino acid variant (*e.g.*, an insertion, substitution and/or deletion derivative) of the *gp35* sequence shown in Figure

10 2. In a specific embodiment, such insertion, substitution and/or deletion occur outside of amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79 or 81-93 depicted in Figure 2.

The invention also relates to gp35 analogs.

The gp35 fragment, amino acid variant or analog of the invention is capable of

15 displaying one or more functional activities associated with a full-length native gp35 protein. Such functional activities include, but are not limited to, antigenicity, *i.e.*, the ability to bind to an anti-gp35 antibody, immunogenicity, *i.e.*, the ability to generate an antibody which is capable of binding a gp35 protein; the ability to bind (or compete with gp35 for binding) to a ligand for gp35; and the ability to multimerize with P36 and/or P34.

20 For an example of the latter, a functional ability of the gp35 protein is the ability of gp35 or a gp35-P36 oligomer to bind to P34 and/or the ability of gp35 to bind to P36.

In a specific embodiment, the invention provides gp35 fragments or variants that comprise at least a functionally active portion of the gp35 sequence shown in Figure 2 from amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-79, or 81-93.

25 In a specific embodiment, the invention provides derivatives (including fragments) or analogs of a gp35 protein consisting of at least 8 contiguous amino acids, or of at least 15 contiguous amino acids, or of at least 20 contiguous amino acids, of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 24. In a preferred embodiment, this derivative or analog is able to be bound by an antibody directed

30 against a gp35 protein. In another preferred embodiment, the derivative or analog specifically binds the P34 homooligomer. Nucleic acids encoding such derivatives or analogs are also within the scope of the invention.

The invention further provides derivatives, fragments or analogs of a gp35 protein consisting of at least 40, 45, 50, 60, or 70 contiguous amino acid residues of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 100. In a specific embodiment, this gp35 derivative, fragment or analog lacks amino acid

5   residues 93 to 372.,

The invention further relates to fragments (and derivatives and analogs thereof) of gp35 which comprise one or more functional domains of a gp35 protein, e.g., the P36 or P34 binding domain, and/or retain the antigenicity of a gp35 protein (i.e., are able to be bound by an anti-gp35 antibody). In specific embodiments, the fragments lack at least 10,

10   20, 30 or 40 contiguous amino acids of the sequence shown in Figure 2.

The invention also relates to gp35 proteins, derivatives and analogs in which internal peptide sequences are deleted without affecting the ability of gp35 to associate with its natural tail fiber partners P36 and/or P34. In a specific embodiment, the deletion occurs of contiguous amino acids selected from among amino acids 100-273. In other embodiments,

15   gp35 is modified so that it interacts only with other modified, and not native, tail fiber partners; exhibit thermolabile interactions with its partners; or contains, or is conjugated to, additional functional groups that enables it to interact with heterologous binding moieties.

The gp35 protein, or derivatives or analogs thereof, described herein, may be used for the production of anti-gp35 antibodies, which antibodies may be used in immunoassays

20   for the detection or measurement of gp35 protein.

The present invention also relates to a gp35 protein, derivative or analog that is modified in the domain that which forms an angle joint, to form an average angle that is different from the natural average angle of 137° (±7°) or 156° (±12°).

The present invention further relates to methods of production of the gp35 proteins,

25   derivatives and analogs, such as, for example, by recombinant means.

The present invention additionally provides for nanostructures comprising native or modified gp35 and native or modified bacteriophage tail fiber proteins. The nanostructures may be one-dimensional rods, two-dimensional polygons or open or closed sheets, or three-dimensional open cages or closed solids. The gp35 protein may be modified in

30   various ways to form novel structures with different properties for use as described in Section 5.8.

5.1. THE gp35 CODING SEQUENCES

gp35 DNA sequences and sequences complementary thereto are gp35 nucleic acids provided by the present invention. Sequences hybridizable thereto, are also provided. Nucleic acids comprising gp35 DNA or RNA sequences are also provided; in various
5 embodiments, at least 850, 880, 920, 960, or 1000 contiguous nucleotides of the gp35 sequence in Figure 2, are in the nucleic acid. Also included within the scope of the present invention are nucleic acids comprising gp35 DNA having the sequence depicted in Figure 2 (SEQ ID NO:2), or its corresponding RNA, which do not encode other bacteriophage T4 tail fiber proteins or functionally active portions thereof.

10 Nucleic acids can be single-stranded or double-stranded. In specific embodiments, isolated nucleic acids are provided that comprise at least 150, 175, 200, 225, 250, 275, or 285 contiguous nucleotides of nucleotides 1 to 285 in Figure 2.

In specific embodiments, the nucleic acids of the invention comprise the nucleotide sequences shown in Figure 2 that encode amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17,
15 57-93, 57-64, 66-79, or 81-93 of Figure 2.

In another embodiment, the nucleic acids comprise nucleotide numbers 1 to 1,116 of Figure 2.

The gp35 nucleotide sequences of the invention, preferably do not contain in contiguous linkage sequences of a bacteriophage T4 genome that are naturally in contiguous
20 linkage flanking the gp35 sequences (i.e., 5' or 3' to the gp35 gene). For example, the gp35 nucleotide sequences can be contiguous with non-bacteriophage T4 nucleotide sequences of at least 10 nucleotides.

In a specific embodiment, the invention provides an isolated nucleic acid comprising a nucleotide sequence encoding a gp35 protein having the amino acid sequence depicted in
25 Figure 2 (SEQ ID NO:2), operably linked to a heterologous promoter. By "heterologous promoter" is meant a promoter that is not the native T4 promoter that is operably linked to the gp35 sequence in the bacteriophage T4 genome. In a specific embodiment, the promoter is not a bacteriophage T4 promoter. In a preferred embodiment, the nucleotide sequence encoding the gp35 protein is that sequence depicted in Figure 2 (SEQ ID NO:1)
30 from nucleotide numbers 1 to 1,116 contiguous to a 3' termination codon.

In other specific embodiments, nucleic acids contain at least 850, 880, 920, 960, or 1000 contiguous nucleotides of a *gp35* DNA sequence operably linked to a promoter that is not a bacteriophage promoter (i.e., a heterologous promoter).

In a specific embodiment, the nucleic acid further comprises nucleotide sequences

5   encoding other bacteriophage T4 proteins selected from the group consisting of gp36 and gp37, and optionally the chaperon protein gp57, operably linked to the same or a different promoter. Preferably, native intergenic regions between the other bacteriophage T4 proteins are omitted.

The invention also provides single-stranded oligonucleotides for use as primers in

10   PCR that amplify a *gp35* gene or *gp35* sequence-containing fragment, *e.g.*, an oligonucleotide having the sequence of a hybridizable portion (at least ~8 nucleotides) of *gp35*, and another oligonucleotide having the reverse complement of a downstream sequence in the same strand of *gp35*, such that each oligonucleotide primes synthesis in a direction toward the other. In one embodiment, the 5' oligonucleotide corresponds to

15   sequence flanking nucleotides 1-280 of Figure 2. In a specific embodiment, the 5' primer comprises a sequence upstream of nucleotide number 1 in Figure 2 and/or also comprises a nucleotide sequence shown in Figure 2 encoding an amino-terminal portion (*i.e.* at least the N-terminal amino acid) of gp35. In a specific embodiment, the oligonucleotide primers are preferably in the range of 10-35 nucleotides in length. A kit comprising in one or more

20   containers the foregoing primers is also provided.

The full length sequence for *gp35* is depicted in Figure 2 (SEQ ID NO:1), with the coding region thereof spanning nucleotide numbers 1 to 1,116. Sequence analysis of the nucleotide sequence of *gp35* of Figure 2 reveals an open reading frame of 1,116 nucleotides, encoding a protein of 372 amino acids (SEQ ID NO:2).

25   In accordance with the present invention, any polynucleotide sequence which encodes the amino acid sequence of a *gp35* product can be used to generate recombinant molecules which direct the expression of gp35. Included within the scope of the present invention are nucleic acids consisting of at least 8 nucleotides that are useful as probes or primers (*i.e.*, a hybridizable portion) in the detection or amplification of *gp35*. In a

30   preferred embodiment, these probes or primers have a contiguous sequence contained in nucleotides 1 to 279 of Figure 2. The invention also relates to nucleic acid sequences hybridizable or complementary to the foregoing sequences or equivalent to the foregoing

sequences in that the equivalent nucleic acid sequences also encode a protein product displaying gp35 functional activity.

Nucleic acids encoding fragments and derivatives of gp35 are additionally described *infra*.

5   The invention also relates to nucleic acids hybridizable to or complementary to the above-described nucleic acids comprising *gp35* sequences. In specific aspects, nucleic acids are provided which comprise a sequence absolutely complementary to at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of a *gp35* gene, or, in particular, those portions encoding amino acid numbers 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-64, 66-

10 79, and/or 81-93 of Figure 2. In a specific embodiment, a nucleic acid which is hybridizable to a *gp35* nucleic acid, or to a nucleic acid encoding a gp35 derivative, under conditions of low stringency is provided. By way of example and not limitation, procedures using such conditions of low stringency are as follows (see also Shilo and Weinberg, 1981, *Proc. Natl. Acad. Sci. USA* 78:6789-6792): Filters containing DNA are pretreated for 6 h at

15 40°C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 μg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 μg/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in

20 hybridization mixture for 18-20 h at 40°C, and then washed for 1.5 h at 55°C in a solution containing 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 h at 60°C. Filters are blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and reexposed to film. Other conditions of low stringency which may be

25 used are well known in the art (*e.g.*, as employed for cross-species hybridizations).

In another specific embodiment, a nucleic acid which is hybridizable to a *gp35* nucleic acid under conditions of high stringency is provided (see *infra*).

The DNA may be obtained by standard procedures known in the art from, for example, by chemical synthesis or by the cloning the DNA, or fragments thereof, purified

30 from a desired cell or phage. (*See*, for example, Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, *DNA Cloning: A Practical Approach*, MRL Press,

Ltd., Oxford, U.K. Vol. I, II.). Whatever the source, the gene should be molecularly cloned into a suitable vector for propagation of the gene.

In the molecular cloning of the gene from DNA preparations, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at

5   specific sites using various restriction enzymes. The linear DNA fragments can then be separated according to size by standard techniques, including, but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography. See, for example, Innis et al., 1990, PCR protocols: A Guide to Methods and Applications, Academic Press, San Diego, California; Dieffenbach et al., 1995, PCR primer, A Laboratory Manual, Cold

10 Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the desired gene may be accomplished in a number of ways. For example, a *gp35* gene of the present invention or its specific RNA, or a fragment thereof, such as a probe or primer, may be isolated and labeled and then used in hybridization assays to detect

15 a generated *gp35* sequence (Benton, W. and Davis, R., 1977, *Science* 196:180; Grunstein, M., and Hogness, D., 1975, *Proc. Natl. Acad. Sci. USA* 72:3961). Those DNA fragments sharing substantial sequence homology to the probe will hybridize, e.g., under high stringency conditions. By way of example, the phrase "high stringency conditions" as used herein refers to those hybridizing conditions that (1) employ low ionic strength and high

20 temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at 50°C; (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075

25 M sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS.

It is also possible to identify the appropriate fragment by restriction enzyme digestion(s) and comparison of fragment sizes with those expected according to a known

30 restriction map. Further selection can be carried out on the basis of the properties of the gene. Alternatively, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example,

DNA clones which hybrid-select the proper mRNAs, can be selected which produce a protein that has similar or identical electrophoretic migration, isolectric focusing behavior, proteolytic digestion maps, binding activity or antigenic properties as known for gp35. Alternatively, the gp35 protein may be identified by binding of labeled antibody to the

5   putatively gp35 expressing clones, *e.g.*, in an ELISA (enzyme-linked immunosorbent assay)-type procedure.

gp35 sequence can also be identified by mRNA selection by nucleic acid hybridization followed by *in vitro* translation. In this procedure, fragments are used to isolate complementary mRNAS by hybridization. Such DNA fragments may represent

10   available, purified gp35 DNA of a naturally occurring or modified *gp35* gene. Immunoprecipitation analysis or functional assays of the *in vitro* translation products of the isolated products of the isolated mRNAS identifies the mRNA and, therefore, the complementary DNA fragments that contain the desired sequences. Radiolabelled RNA or DNA may be used as a probe to identify the *gp35* DNA fragments from among other DNA

15   fragments.

Alternatives to isolating *gp35* DNA include, but are not limited to, chemically synthesizing the gene sequence itself from a known sequence. Other methods are known to those of skill in the art and are within the scope of the invention.

The identified and isolated DNA can then be inserted into an appropriate cloning

20   vector. A large number of vector-host systems known in the art may be used. Possible vectors include, but are not limited to, plasmids or modified viruses, but the vector system must be compatible with the host cell used. Such vectors include, but are not limited to, bacteriophages such as lambda or T4 derivatives, or plasmids such as PBR322 or pUC plasmid derivatives or the Bluescript vector (Stratagene). The insertion into a cloning

25   vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated

30   linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. In an alternative method, the cleaved vector and *gp35* sequence may be modified by homopolymeric tailing. Recombinant molecules can be

techniques, cleaved from the resin, and purified by preparative high performance liquid chromatography (*e.g.*, *see* Creighton, 1983, Proteins Structures And Molecular Principles, W.H. Freeman and Co., N.Y. pp. 50-60). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure;

5   *see* Creighton, 1983, Proteins, Structures and Molecular Principles, W.H. Freeman and Co., N.Y., pp. 34-49).

In order to express a biologically active gp35 protein or functional equivalent thereof, a polynucleotide sequence encoding a gp35 protein, or derivative or analog thereof, is inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary

10   elements for the transcription and translation of the inserted coding sequence. The *gp35* gene products as well as host cells or cell lines transfected or transformed with recombinant gp35 expression vectors can be used for a variety of purposes. These include, but are not limited to, producing gp35 protein for use as an immunogen for generating antibodies (*i.e.*, monoclonal or polyclonal) that immunospecifically bind a gp35 protein and providing gp35

15   protein building blocks for nanostructures containing bacteriophage tail fiber proteins or protein derivatives.

5.2.1. EXPRESSION SYSTEMS

Methods known to those skilled in the art can be used to construct expression

20   vectors containing a gp35 coding sequence of interest (native, modified, or recombined) and appropriate transcriptional/translational control signals. These expression vectors typically contain selectable marker genes (usually conferring antibiotic resistance to transformed bacteria), sequences that allow replication of the plasmid to high copy number in *E. coli*, and a multiple cloning site immediately downstream of an inducible promoter and ribosome

25   binding site. Methods of constructing expression vectors containing a gp35 coding sequence include *in vitro* recombinant DNA techniques and synthetic techniques. *See*, for example, the techniques described in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory, N.Y. and Ausubel et al., 1989, Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley

30   Interscience, N.Y.

A variety of host-expression vector systems may be utilized to express a gp35 coding sequence. These systems are preferably bacteria transformed with recombinant

- 16 -

bacteriophage DNA or plasmid DNA expression vectors containing a gp35 coding sequence, but also include, but are not limited to, yeast transformed with recombinant yeast expression vectors containing an gp35 coding sequence; insect cell systems infected with recombinant virus expression vectors (*e.g.*, baculovirus) containing an gp35 coding

5 sequence; plant cell systems infected with recombinant virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (*e.g.*, Ti plasmid) containing a gp35 coding sequence; or animal cell systems. The expression elements of these systems vary in their strength and specificities. Depending on the host/vector system utilized, any of a number of

10 suitable transcription and translation elements, including constitutive and inducible promoters, may be used in the expression vector. For example, when cloning in bacterial systems, inducible promoters such as Pl of bacteriophage λ, ptrp, ptac (ptrp-lac hybrid promoter) and the like may be used. A preferred promoter is plac (with a laci^q on the vector to reduce background expression). A second preferred promoter is pT7ϕ10, which is

15 specific to T7 RNA polymerase and is not recognized by *E. coli* RNA polymerase.

Examples of other host systems, include, but are not limited to; cloning in insect cell systems using promoters such as the baculovirus polyhedrin promoter; cloning in plant cell systems using promoters derived from the genome of plant cells (*e.g.*, heat shock promoters; the promoter for the small subunit of RUBISCO; the promoter for the chlorophyll a/b

20 binding protein) or from plant viruses (*e.g.*, the 35S RNA promoter of CaMV; the coat protein promoter of TMV); cloning in mammalian cell systems using promoters derived from the genome of mammalian cells (*e.g.*, metallothionein promoter) or from mammalian viruses (*e.g.*, the adenovirus late promoter; the vaccinia virus 7.5 K promoter); and generating cell lines that contain multiple copies of a gp35 DNA, SV40-, BPV- and EBV-

25 based vectors may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the gp35 protein derivative or analog expressed. For example, when large quantities of gp35 protein, derivative or analog are to be produced for the generation of antibodies, vectors which direct the expression of high

30 levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited to, the *E. coli* expression vector pUR278 (Ruther et al., 1983, *EMBO J.* 2:1791), in which the gp35 coding sequence may be ligated into the vector in

frame with the *lacZ* coding region so that a hybrid AS-lacZ protein is produced; PIN vectors (Inouye & Inouye, 1985, *Nucleic acids Res.* 13:3101-3109; Van Heeke & Schuster, 1989, *J. Biol. Chem.* 264:5503-5509); and the like. PGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST) (Smith and

5   Johnson, 1988, *Gene* 7:31-40). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The PGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety.

10   In yeast, a number of vectors containing constitutive or inducible promoters may be used. For a review *see*, Current Protocols in Molecular Biology, Vol. 2, 1988, Ed. Ausubel et al., Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant et al., 1987, Expression and Secretion Vectors for Yeast, *in* Methods in Enzymology, Ed. Wu & Grossman, 1987, Acad. Press, N.Y. 153:516-544; Glover, 1986, DNA Cloning, Vol. II, IRL Press, Wash.,

15   D.C., Ch. 3; and Bitter, 1987, Heterologous Gene Expression in Yeast, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y. 152:673-684; and The Molecular Biology of the Yeast Saccharomyces, 1982, Eds. Strathern et al., Cold Spring Harbor Press, Vols. I and II.

         In cases where plant expression vectors are used, the expression of a gp35 coding

20   sequence may be driven by any of a number of promoters. For example, viral promoters such as the 35S RNA and 19S RNA promoters of CaMV (Brisson et al., 1984, Nature 310:511-514), or the coat protein promoter of TMV (Takamatsu et al., 1987, EMBO J. 6:307-311) may be used; alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi et al., 1984, EMBO J. 3:1671-1680; Broglie et al., 1984, Science

25   224:838-843); or heat shock promoters, *e.g.*, soybean hsp17.5-E or hsp17.3-B (Gurley et al., 1986, Mol. Cell. Biol. 6:559-565) may be used. These constructs can be introduced into plant cells using Ti plasmids, Ri plasmids, plant virus vectors, direct DNA transformation, microinjection, electroporation, etc. For reviews of such techniques *see*, for example, Weissbach & Weissbach, 1988, Methods for Plant Molecular Biology, Academic Press,

30   NY, Section VIII, pp. 421-463; and Grierson & Corey, 1988, Plant Molecular Biology, 2d Ed., Blackie, London, Ch. 7-9.

●   ●

An alternative expression system which could be used to express a gp35 gene is an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. A gp35 coding sequence may be cloned into non-essential regions (for

5 example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example, the polyhedrin promoter). Successful insertion of a gp35 coding sequence will result in inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in

10 which the inserted gene is expressed. (*e.g.*, see Smith et al., 1983, J. Virol. 46:584; Smith, U.S. Patent No. 4,215,051).

In mammalian host cells, a number of viral based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a gp35 coding sequence may be ligated to an adenovirus transcription/translation control complex, *e.g.*, the

15 late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) will result in a recombinant virus that is viable and capable of expressing gp35 in infected hosts. (*e.g.*, see Logan & Shenk, 1984, Proc. Natl. Acad. Sci. USA 81:3655-3659). Alternatively, the vaccinia 7.5 K promoter may be

20 used. (*See, e.g.*, Mackett et al., 1982, Proc. Natl. Acad. Sci. USA 79:7415-7419; Mackett et al., 1984, J. Virol. 49:857-864; Panicali et al., 1982, Proc. Natl. Acad. Sci. USA 79:4927-4931).

Other examples of commercially available vectors suitable for use in a bacteria host include, but are not limited to, the PET system (Novagen, Inc., Madison, WI) and

25 Superlinker vectors PSE280 and PSE380 (Invitrogen, San Diego, CA).

Specific initiation signals may also be required for efficient translation of an inserted gp35 coding sequence. These signals include the ATG initiation codon and adjacent sequences. In cases where an entire *gp35* gene, including its own initiation codon and adjacent sequences, is inserted into the appropriate expression vector, no additional

30 translational control signals may be needed. However, in cases where only a portion of a gp35 coding sequence is inserted, lacking the 5' end, exogenous translational control signals, including the ATG initiation codon, must be provided. Furthermore, the initiation

- 19 -

codon must be in phase with the reading frame of a gp35 coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer

5   elements, transcription terminators, etc. (*see* Bittner et al., 1987, *Methods in Enzymol.* 153:516-544).

In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications and processing (*e.g.*, cleavage) of protein products may be

10   important for the function of the protein. Different host cells have characteristic and specific mechanisms for post-transcriptional and post-translational processing and modification. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript,

15   and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERO, BHK, HeLa, COS, MDCK, 293, WI38, etc.

Preferred hosts for producing the proteins of the present invention are *E. coli* strains BL21 (DE3) and BL21 (DE/plys5) (NoVagen, Madison, Wisconsin).

For long-term, high-yield production of recombinant proteins, stable expression is

20   preferred. For example, cell lines which stably express a gp35 protein, derivative or analog may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with *gp35* DNA controlled by appropriate expression control elements (*e.g.*, bacterial promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), a selectable marker, and flanked by sequences that

25   promote homologous recombination. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows for the stable integration of the plasmid into host chromosomes. This method may advantageously be used to engineer bacterial strains

30   which express a gp35 protein.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler et al., 1977, Cell 11:223), hypoxanthine-guanine

phosphoribosyltransferase (Szybalska & Szybalski, 1962, Proc. Natl. Acad. Sci. USA 48:2026), and adenine phosphoribosyltransferase (Lowy et al., 1980, Cell 22:817) genes can be employed in tk-, hgprt- or aprt- cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler et

5 al., 1980, Natl. Acad. Sci. USA 77:3567; O'Hare et al., 1981, Proc. Natl. Acad. Sci. USA 78:1527); gpt, which confers resistance to mycophenolic acid (Mulligan & Berg, 1981, Proc. Natl. Acad. Sci. USA 78:2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin et al., 1981, J. Mol. Biol. 150:1); and hygro, which confers resistance to hygromycin (Santerre et al., 1984, Gene 30:147). Recently, additional

10 selectable genes have been described, namely trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman & Mulligan, 1988, Proc. Natl. Acad. Sci. USA 85:8047); and ODC (ornithine decarboxylase) which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue, L., 1987, In: Current

15 Communications in Molecular Biology, Cold Spring Harbor Laboratory, Ed.).

The present invention provides a method for producing a recombinant gp35 protein, derivative or analog comprising culturing a host cell transformed with a recombinant expression vector encoding a gp35 protein, derivative or analog, such that the gp35 protein, derivative or analog is expressed by the cell and recovering the expressed gp35 protein,

20 derivative or analog.

5.2.2. IDENTIFICATION OF TRANSFECTANTS OR TRANSFORMANTS THAT EXPRESS gp35

The host cells which contain the coding sequence and which express the gp35 product or functionally active derivatives or analogs thereof may be identified by at least

25 four general approaches; (a) DNA-DNA or DNA-RNA hybridization; (b) the presence or absence of "marker" gene functions; (c) assessing the level of transcription as measured by the expression of gp35 mRNA transcripts in the host cell; and (d) detection of the gene product as measured by immunoassay or by its biological activity.

In the first approach, the presence of the gp35 coding sequence inserted in the

30 expression vector can be detected by DNA-DNA or DNA-RNA hybridization using probes

- 21 -

comprising nucleotide sequences that are homologous to the gp35 coding sequence, respectively, or derivatives (e.g., fragments) or analogs thereof.

In the second approach, the recombinant expression vector/host system can be identified and selected based upon the presence or absence of certain "marker" gene

5 functions (e.g., resistance to antibiotics). For example, if the gp35 coding sequence is inserted within a marker gene sequence of the vector, recombinant cells containing the gp35 coding sequence can be identified by the absence of the marker gene function. Alternatively, a marker gene can be placed in tandem with a *gp35* coding sequence under the control of the same or different promoter used to control the expression of the gp35

10 coding sequence. Expression of the marker in response to induction or selection indicates expression of the gp35 coding sequence.

In the third approach, transcriptional activity of *gp35* can be assessed by hybridization assays. For example, RNA can be isolated and analyzed by Northern blot using a probe having sequence homology to a gp35 coding sequence or transcribed

15 noncoding sequence or particular portions thereof. Alternatively, total nucleic acid of the host cell may be extracted and quantitatively assayed for hybridization to such probes.

In the fourth approach, the levels of a gp35 protein, derivative or analog product can be assessed immunologically, for example by Western blots, immunoassays such as radioimmuno-precipitation, enzyme-linked immunoassays and the like.

20

5.3. PURIFICATION OF THE EXPRESSED GENE PRODUCT

Once a recombinant which expresses the *gp35* gene sequence is identified, the gene product can be analyzed. This is achieved by assays based on the physical or functional properties of the product, including radioactive labelling of the product followed by analysis

25 by gel electrophoresis, immunoassay, or other detection methods known to those of skill in the art.

Once the gp35 protein is identified, it may be isolated and purified by standard methods including chromatography (e.g., ion exchange, affinity, and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique

30 for the purification of proteins. The functional properties may be evaluated using any suitable assay.

Alternatively, once a gp35 protein produced by a recombinant is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant. As a result, the protein can be synthesized by standard chemical methods known in the art (*e.g.*, *see* Hunkapiller et al., 1984, *Nature*

5   310:105-111).

In a specific embodiment, the invention relates to a purified gp35 protein that is not contained in a gel suitable for electrophoresis. In a preferred embodiment, the purified gp35 protein is not denatured.

In another specific embodiment, the invention relates to a composition containing at

10   least 1, 10, 50, 100 or 500 nanogram(s), 1, 10, 50, 100 or 500 microgram(s), or 1, 10, 50, 100 or 500 milligram(s), of purified non-denatured gp35 protein. In a preferred embodiment, this composition is not a gel suitable for electrophoresis.

In a specific embodiment of the present invention, such gp35 proteins, whether produced by recombinant DNA techniques or by chemical synthetic methods include, but

15   are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence substantially as depicted in Figure 2 (SEQ ID NO:2), as well as fragments and other derivatives, and analogs thereof.

5.4. GENERATION OF ANTIBODIES TO gp35

20   According to the invention, gp35 protein, its derivatives (e.g., fragments), or analogs thereof, may be used as an immunogen to generate antibodies which recognize such an immunogen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library.

Various procedures known in the art may be used for the production of polyclonal

25   antibodies to a gp35 protein or derivative or analog.

In one embodiment, by way of example, gp35 specific antisera is prepared according to procedures as described by Edgar (1965, *Genetics* 52: 1187) and Ward (1970, *J. Mol. Biol.* 54:15). Briefly, whole T4 bacteriophage are used as an immunogen; the resulting antiserum is then adsorbed with tail-less phage particles, thus removing all antibodies

30   except those directed against the tail fiber proteins. In a subsequent step, different aliquots of the antiserum are adsorbed individually with extracts that each lack a particular tail fiber protein. For example, if an extract containing only tail fiber components gp34, gp36 and

gp37 (derived from a cell infected with a mutant T4 that does not produce gp35) is used for adsorption, the resulting antiserum will recognize only mature gp35 and dimerized gp35-P36 or gp35-P34. In an alternative embodiment, antibody is raised against purified tail fiber halves, *e.g.*, gp35-gp36-gp37. According to this embodiment, anti gp35-gp36-gp37 is

5   then adsorbed with gp36-gp37 to produce anti-gp35. In another embodiment, anti-gp35 is produced directly using purified gp35 proteins, derivatives or analogs thereof, as an immunogen. In another embodiment, monoclonal antibodies are generated against a gp35 protein sequence or analog thereof using techniques known in the art.

For the production of antibody, various host animals can be immunized by injection

10   with the native gp35 protein, or a synthetic version, derivative (*e.g.*, fragment) or analog thereof, including, but not limited to, rabbits, mice, rats, etc. Various adjuvants may be used to increase the immunological response, depending on the host species, and including, but not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions,

15   peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, BCG (bacille Calmette-Guerin) and corynebacterium parvum.

For preparation of monoclonal antibodies directed toward a gp35 protein sequence or analog thereof, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique

20   originally developed by Kohler and Milstein (1975, *Nature* 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, *Immunology Today* 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96). In an additional embodiment of the invention, monoclonal antibodies can be

25   produced in germ-free animals utilizing recent technology (PCT/US90/02545).

According to the invention, techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce gp35-specific single chain antibodies. An additional embodiment of the invention utilizes the techniques described for the construction of Fab expression libraries (Huse et al., 1989, *Science* 246:1275-1281) to

30   allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for gp35 proteins, derivatives, or analogs.

- 24 -

In one embodiment, a molecule comprising a fragment of the gp35 protein is used as an immunogen. In a preferred embodiment, the fragment used as the immunogen has a sequence that is all or a portion of amino acid residues 1 to 93, and lacks amino acid residues 94 to 373 in Figure 2. Since hydrophilic regions are believed most likely to

5　contain antigenic determinants, a peptide corresponding to or containing a hydrophilic portion of a gp35 protein is preferably used as immunogen.

Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include, but are not limited to: the F(ab′)$_2$ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab′

10　fragments which can be generated by reducing the disulfide bridges of the F(ab′)$_2$ fragment, and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent.

In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art, e.g., ELISA (enzyme-linked immunosorbent

15　assay). For example, to select antibodies which recognize a specific domain of a gp35 protein, one may assay generated hybridomas for a product which binds to a gp35 fragment containing such domain.

The foregoing antibodies can be used in methods known in the art relating to the localization and activity of the protein sequences of the invention, e.g., for imaging these

20　proteins, measuring levels thereof, in diagnostic methods, etc.

A non-limiting method by which anti-gp35 may also be used to detect gp35 tail fiber proteins, derivatives or analogs, involves screening for bacterial colonies expressing proteins, derivatives or analogs by directly transferring the colonies, or, alternatively, samples of lysed or unlysed cultures, to nitrocellulose filters, lysing the bacterial cells on

25　the filter if necessary, and incubating with specific antibodies. Formation of immune complexes may then be detected by methods widely used in the art (e.g., secondary antibody conjugated to a chromogenic enzyme or radiolabelled Staphylococcal Protein A.). This method is particularly useful to screen large numbers of colonies. In an alternative method, bacterial cells expressing the protein, derivative, or analog of interest are first

30　metabolically labelled with ^{35}S-methionine, followed by preparation of extracts and incubation with antiserum. The immune complexes may then be recovered by incubation with immobilized Protein A followed by centrifugation and resolution by SDS-PAGE.

●   ●

5.5. STRUCTURE OF THE *gp35* GENE AND PROTEIN

The structure of the *gp35* gene and protein can be analyzed by any of various methods known in the art. Representative methods are set forth below.

5
5.5.1. GENETIC ANALYSIS

The cloned DNA corresponding to *gp35* can be analyzed by methods including, but not limited to, Southern hybridization (Southern, E.M., 1975, *J. Mol. Biol.* 98:503-517), Northern hybridization (*see, e.g.,* Freeman et al., 1983, *Proc. Natl. Acad. Sci. USA* 80:4094-4098), restriction endonuclease mapping (Maniatis, T., 1982, *Molecular Cloning, A*

10 *Laboratory*, Cold Spring Harbor, New York), and DNA sequence analysis. Polymerase chain reaction (PCR; U.S. Patent Nos. 4,683,202, 4,683,195, and 4,889,818; Gyllenstein et al., 1988, *Proc. Natl. Acad. Sci. USA* 85:7652-7656; Ochman et al., 1988, *Genetics* 120:621-623; Loh et al., 1989, *Science* 243:217-220) followed by Southern hybridization with a *gp35*-specific probe. Northern hybridization analysis can be used to determine the

15 expression levels of *gp35*. The stringency of the hybridization conditions for both Southern and Northern hybridization, or dot blots, can be manipulated to ensure detection of nucleic acids with the desired degree of relatedness to the specific *gp35* probe used.

Restriction endonuclease mapping can be used to roughly determine the genetic structure of *gp35*. Restriction maps derived by restriction endonuclease cleavage can be

20 confirmed by DNA sequence analysis.

DNA sequence analysis can be performed by any techniques known in the art, including, but not limited to, the method of Maxam and Gilbert (1980, *Meth. Enzymol.* 65:499-560), the Sanger dideoxy method (Sanger et al., 1977, *Proc. Natl. Acad. Sci. USA* 74:5463), the use of T7 DNA polymerase (Tabor and Richardson, U.S. Patent No.

25 4,795,699), or use of an automated DNA sequenator (*e.g.,* Applied Biosystems, Foster City, CA). The nucleotide sequence of a representative *gp35* gene comprises the sequence substantially as depicted in Figure 2 (SEQ ID NO:1), and described in Section 6, *infra*.

30
5.5.2. PROTEIN ANALYSIS

The amino acid sequence of a gp35 protein, derivative, fragment or analog can be derived by deduction from the DNA sequence, or alternatively, by direct sequencing of the

- 26 -

protein, *e.g.*, with an automated amino acid sequencer. The amino acid sequence of a representative gp35 protein comprises the sequence substantially as depicted in Figure 2 (SEQ ID NO:2), and detailed in Section 6, *infra*, with the representative protein that is shown by amino acid numbers 1-372.

5   The gp35 protein sequence can be further characterized by a hydrophilicity analysis (Hopp, T., and Woods, K., 1981, *Proc. Natl. Acad. Sci. USA* 78:3824). A hydrophilicity profile can be used to identify the hydrophobic and hydrophilic regions of the gp35 protein and the corresponding regions of the DNA sequence which encode such regions. Hydrophilic regions are predicted to be antigenic/immunogenic.

10   Secondary structural analysis (Chou, P., and Fasman, G., 1974, *Biochemistry* 13:222) can also be done, to identify regions of the gp35 protein that assume specific secondary structures.

Manipulation, translation, and secondary structure prediction, as well as open reading frame prediction and plotting, can also be accomplished using computer software

15   programs available in the art.

Other methods of structural analysis can also be employed. These include, but are not limited to, X-ray crystallography (Engstom, A., 1974, *Biochem. Exp. Biol.* 11:7-13) and computer modeling (Fletterick, R., and Zoller, M. (eds.), 1986, Computer Graphics and Molecular Modeling, in *Current Communications in Molecular Biology*, Cold Spring

20   Harbor Laboratory, Cold Spring Harbor, New York).

5.6. gp35 PROTEINS, DERIVATIVES AND ANALOGS

The invention further relates to gp35 proteins, derivatives (including, but not limited to, fragments) and analogs of gp35 proteins. Nucleic acids encoding gp35 proteins,

25   derivatives and analogs are also provided. Molecules comprising gp35 proteins, derivatives or analogs are also provided. In one embodiment, the gp35 proteins, derivatives or analogs are encoded by the *gp35* nucleic acids described in Section 5.1 *supra*.

The production and use of derivatives and analogs related to gp35 are within the scope of the present invention. In a specific embodiment, the derivative or analog is

30   functionally active, *i.e.*, capable of exhibiting one or more functional activities associated with a full-length, wild-type gp35 protein. As one example, such derivatives or analogs which have the desired immunogenicity or antigenicity can be used, for example, in

- 27 -

immunoassays, for inhibition of gp35 activity, etc. As another example, such derivatives or analogs which are able to bind bacteriophage T4 tail fiber proteins P36 and/or P34 are provided. Derivatives or analogs that retain a desired gp35 property of interest (*e.g.*, binding to tail fiber proteins), can be used as inhibitors of such property and its

5 physiological correlates. A specific embodiment relates to a gp35 fragment that can be bound by an anti-gp35 antibody. Derivatives or analogs of gp35 can be tested for the desired activity by procedures known in the art, including, but not limited to, the assays described *infra*.

In particular, gp35 derivatives can be made by altering *gp35* sequences by

10 substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding sequences, other DNA sequences which encode substantially the same amino acid sequence as a *gp35* gene may be used in the practice of the present invention. These include, but are not limited to, nucleotide sequences comprising all or portions of *gp35* which are altered by the substitution of different codons

15 that encode a functionally equivalent amino acid residue within the sequence, thus producing a silent change. Likewise, the gp35 derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of a gp35 protein including altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in

20 a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Conservative substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine,

25 valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

In one embodiment, the invention relates to non-native bacteriophage T4 gp35

30 proteins, derivatives or analogs in which only conservative substitutions relative to the sequence in Figure 2 are made.

The invention also relates to non-native molecules encoded by a nucleic acid that is capable of hybridizing to gp35 coding sequence (SEQ ID NO:1), under stringent, moderately stringent, or nonstringent conditions.

In another embodiment, the invention relates to proteins, derivatives or analogs

5   comprising the amino acid sequence depicted in Figure 2 (SEQ ID NO:2) from amino acid residues 1-17, 1-56, 1-78, 1-93, 8-17, 57-93, 57-93, 57-64, 66-79, or 81-93. In another embodiment, these proteins contain only conservative substitutions relative to the sequence in Figure 2.

The invention additionally relates to proteins, derivatives or analogs, comprising an

10   amino acid sequence that has at least 60%, 65%, 70%, 75%, 80%, 85%, or 90% amino acid sequence homology, to bacteriophage T4 gp35 amino acids number 1 to 100 in Figure 2 over a 100 amino acid sequence.

The invention further relates to proteins, derivatives, fragments or analogs comprising an amino acid sequence sharing at least 30%, 35%, 40%, 45%, 50%, 55%, 60%,

15   65%, 70%, 75%, 80%, 85% or 90% homology to amino acids numbers 57 to 93 in Figure 2 over a 36 amino acid sequence.

The invention further provides derivatives, fragments or analogs of a gp35 protein consisting of at least 8, 15, or 20 contiguous amino acids of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 24. In one

20   embodiment, the derivative, fragment or analog is not native and contains only conservative substitutions relative to the sequence in Figure 2. In a preferred embodiment, the derivative or analog additionally displays one or more functional activities of a gp35 protein. In another preferred embodiment, the derivative, fragment or analog specifically binds P34 and/or P36. In another preferred embodiment, the derivative or analog is able to be bound

25   by an antibody directed against a gp35 protein in which only conservative substitutions relative to the sequence in Figure 2 are made.

The invention also provides derivatives or analogs of a gp35 protein consisting of at least 40, 45, 50, 60, or 70 contiguous amino acid residues of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids numbers 1 to 100. In a specific

30   embodiment, this derivative lacks amino acid residues 93 to 372.

Tail fiber assembly takes place in a predetermined, ordered interaction of specific bacteriophage protein subunits. The angled joint of the tail fiber is formed by the two step

process in which first, the N-terminus of P36 attaches to the carboxy terminal region of a gp35 monomer and second the N-terminal region of gp35-P36 oligomer then attaches to the C-terminus of P34. In a specific embodiment, a gp35 mutant/derivative or analog is provided in which the interaction of the gp35 derivative or analog with P34 is independent

5 of the gp35 first interacting with P36.

In another embodiment of the invention, gp35 derivatives or analogs form average angles with other tail fiber proteins that are different from the native angle of 137° or 158°. In specific embodiments, the angle joint forms average angles of less than about 90°, 100°, 110°, 120°, or 125°, or more than about 145°, 155°, 165°, under conditions wherein the

10 wild-type gp35 protein forms an angle of 137° when combined with P36-P37 and P34 dimers or trimers. In other embodiments, the angle joint of gp35 proteins, derivatives or analogs exhibit more or less flexibility than the native polypeptide. gp35 sequence variants can be screened for the ability to form such an angle.

Thermolabile structures have many uses in nanostructure construction, such as, for

15 example, initiation of structure assembly at low temperature and subsequent inactivation of and separation from the initiator at high temperature. In one embodiment of the invention, gp35 derivatives and analogs exhibit thermolabile interactions with cognate partners. For example, in one embodiment the interaction of a gp35 derivative with a P36 protein oligomer of bacteriophage T4 is unstable at a temperature of about 40°C, 45°C, 50°C,

20 55°C or 60°C (see Section 7). In another embodiment, the interaction of a gp35 derivative with a P34 protein oligomer of bacteriophage T4 is unstable at a temperature of about 40°C, 45°C, 50°C, 55°C or 60°C (see Section 7). In a specific embodiment, the thermolabile interaction between gp35 and cognate partners is reversible, thereby permitting reattachment of the appropriate termini when the lower temperature is restored, in another

25 specific embodiment, this interaction is irreversible.

In another specific embodiment, the gp35 derivative or analog interacts with only mutant cognate partners (e.g., see Section 7).

In another embodiment, gp35 derivatives or analogs contain a mutant amino acid sequence, or are conjugated to a fixed group, that confers specific binding properties on the

30 entire molecule, e.g., sequences derived from avidin that recognize biotin, sequences derived from immunoglobulin heavy chain that recognize Staphylococcal A protein, sequences derived from the Fab portion of the heavy chain of monoclonal antibodies to

- 30 -

which their respective Fab light chain counterparts could attach and form an antigen-binding site, immunoactive sequences that recognize specific antibodies, or sequences that bind specific metal ions (e.g., divalent metal ions). These ligands may be immobilized to facilitate purification and/or assembly.

5   In a specific embodiment of the invention, proteins consisting of or comprising a fragment of a gp35 protein consisting of at least 8 (continuous) amino acids of the gp35 protein sequence depicted in Figure 2 (SEQ ID NO:2) from amino acids number 1 to 24. In other embodiments, the fragment consists of at least 15 or 20 amino acids of the gp35 protein depicted in Figure 2 from amino acids number 1-24.

10   The invention also provides fragments of a gp35 protein consisting of at least 40, 45, 50, 55, 60, or 70 contiguous amino acid residues of the gp35 sequence in Figure 2 (SEQ ID NO:2) from amino acids 1-100. In specific embodiments, such fragments are not larger than 75, 100 or 150 amino acids. In other specific embodiments, such fragments lack amino acid number 93 to 372 in Figure 2. Derivatives or analogs of gp35 include, but are not

15 limited to, those molecules comprising regions that are substantially homologous to gp35 or fragments thereof (e.g., in various embodiments, at least 60% or 70% or 80% or 90% or 95% identity over an amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art) or whose encoding nucleic acid is capable of hybridizing to a gp35 coding sequence,

20 under stringent, moderately stringent, or nonstringent conditions.

     The gp35 derivatives and analogs of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, the cloned gp35 sequence can be modified by any of numerous strategies known in the art (Maniatis, T., 1990, *Molecular Cloning, A Laboratory*

25 *Manual*, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*. In the production of the gene encoding a derivative or analog of gp35, care should be taken to ensure that the modified gene remains within the same translational reading frame as gp35, uninterrupted

30 by translational stop signals, in the gene region where the desired gp35 activity is encoded.

     Additionally, the gp35-encoding nucleic acid sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to

create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used, including, but not limited to, chemical mutagenesis, *in vitro* site-directed mutagenesis (Hutchinson, C., et al., 1978, *J. Biol. Chem* 253:6551), PCR

5 amplification using primers with altered sequences, etc.

Manipulations of the gp35 sequence may also be made at the protein level. Included within the scope of the invention are gp35 protein fragments or other derivatives or analogs which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking

10 groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including, but not limited to, specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH$_4$; acetylation, formylation, oxidation, reduction; etc.

15 In addition, analogs and derivatives of gp35 can be chemically synthesized. For example, a peptide corresponding to a specific portion of a gp35 protein (see Section 5.6.1), or which mediates the desired activity *in vitro*, can be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the gp35 sequence. Non-

20 classical amino acids include, but are not limited to, the D-isomers of the common amino acids, α-amino isobutyric acid, 4-aminobutyric acid, Abu, 2-amino butyric acid, γ-Abu, ε-Ahx, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β-alanine, fluoro-amino

25 acids, designer amino acids such as β-methyl amino acids, Cα-methyl amino acids, Nα-methyl amino acids, and amino acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

In another embodiment, the gp35 derivative is a molecule comprising a region of homology with a gp35 protein. By way of example, in various embodiments, a first protein

30 region can be considered "homologous" to a second protein region when the amino acid sequence of the first region is at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 90%, or 95% identical, when compared to any sequence in the second region of an equal number of

amino acids as the number contained in the first region or when compared to an aligned sequence of the second region that has been aligned by a computer homology program known in the art. For example, a molecule can comprise one or more regions homologous to a gp35 region (see Section 5.6.1) or a full-length gp35 protein.

5   In another embodiment, the gp35 proteins, derivatives, fragments or analogs of the invention are combined with other tail fiber proteins, derivatives, fragments and/or analogs, to form polygons. In a preferred embodiment, a polygon is formed using the gp35 protein, derivative, or analog of the invention in combination with a P36-34 chimer rod unit as described in PCT Publication WO 96/11947, dated April 25, 1996.

10

5.7. ASSAYS OF gp35 PROTEINS, DERIVATIVES AND ANALOGS

The functional activity of gp35 proteins, derivatives and analogs can be assayed by various methods.

15   For example, in one embodiment, where one is assaying for the ability to bind or compete with wild-type gp35 for binding to anti-gp35 antibody, various immunoassays known in the art can be used, including, but not limited to, competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitin reactions, immunodiffusion assays, *in situ* immunoassays (using 20 colloidal gold, enzyme or radioisotope labels, for example), western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. In one embodiment, antibody binding is detected by 25 detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labelled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention.

30   In another embodiment, where a gp35-binding protein is identified (*e.g.*, P34 and P36), the binding can be assayed, by means well-known in the art.

A nonlimiting method by which antibodies specific to gp35 proteins may be used to assay for the ability of gp35 proteins, derivatives or analogs to associate with other tail fiber proteins involves screening for bacterial colonies expressing mature tail fiber proteins by directly transferring the colonies, or, alternatively, samples of lysed or unlysed cultures, to

5   nitrocellulose filters, lysing the bacterial cells on the filter if necessary, and incubating with antibodies specific for gp35 and its binding partner and detecting the formation of immune complexes by methods widely used in the art (e.g., secondary antibody conjugated to a chromogenic enzyme or radiolabelled Staphylococcal Protein A). Another nonlimiting method involves metabolically labelling bacterial cells expressing gp35 with ^{35}S-

10   methionine, preparing and incubating extracts of these cells with gp35 antiserum, recovering immune complexes by incubation with immobilized Protein A followed by centrifugation, and resolving the proteins by SDS-polyacrylamide gel electrophoresis.

A nonlimiting competitive assay for testing whether gp35 derivatives or analogs such as internally deleted tail fiber proteins that do not permit phage infection nonetheless

15   retain the ability to associate with their appropriate partners utilizes an *in vitro*, complementation system which involves mixing a bacterial extract containing the modified gp35 tail fiber protein with a second extract prepared from cells infected with a phage that is a *gp35* null mutant and therefore does not produce gp35. After several hours of incubation, a third extract is added that contains wild-type gp35, and incubation is continued for several

20   additional hours. Finally, the extract is titered for infectious phage particles by infecting *E. coli* and quantifying the phage plaques that result. A modified gp35 protein, derivative or analog that correctly associates with its tail fiber partners is incorporated into tail fibers in a non-functional manner in the first mixture, thereby preventing the incorporation of the wild-type version of the protein after addition of the third extract; the result is a reduction in

25   the titer of the resulting phage sample. By contrast, if the modified gp35 protein, derivative or analog is unable to associate with its binding partner, it will not be incorporated into phage particles in the first mixture and, thus, will not compete with assembly of intact phage particles when the third extract is added; the phage titer should thus be equivalent to that observed when no modified gp35 is added in to the first mixture (a negative control).

30   Assays for testing whether gp35 proteins, derivatives, such as internally deleted proteins, or analogs that do not permit phage infection nonetheless retain the ability to associate with appropriate tail fiber partners can also be performed *in vivo*. These assays

detect the ability of gp35 proteins, derivatives, or analogs to compete with normal phage parts for assembly, thus reducing the burst size of a wild-type phage infecting the same host cell in which gp35 proteins, derivatives, or analogs are recombinantly expressed. Thus, expression from an expression vector encoding the gp35 proteins, derivative, or analogs is

5   induced inside a cell, which cell is then infected by a wild-type phage. Inhibition of wild-type phage production demonstrates the ability of the recombinant gp35 protein, derivative, or analog to associate with the appropriate tail fiber proteins of the phage.

The above-described methods may be used, alone and in combination, in the design and production of different types of modified gp35 tail fiber proteins. For example, a

10   preliminary screen of a large number of bacterial colonies for those expressing properly associated P34-gp35 and/or gp35-P36 complexes will identify positive colonies, which can then be individually tested by *in vitro* complementation.

Other methods will be known to the skilled artisan and are within the scope of the invention.

15

5.8. APPLICATIONS OF NANOMETER STRUCTURE

The gp35 proteins, derivatives, and analogs of the invention have use in the construction of nanostructures. The uses of such nanostructures are manifold and include applications that require highly regular, well-defined arrays of fibers, cages, or solids, which

20   may include specific attachment sites that allow them to associate with other materials.

In one embodiment, a three-dimensional hexagonal array of tubes is used as a molecular sieve or filter, providing regular vertical pores of precise diameter for selective separation of particles by size. Such filters can be used for sterilization of solutions (*i.e.*, to remove microorganisms or viruses), or as a series of molecular-weight cut-off filters. In

25   this case, the protein components of the pores may be modified so as to provide specific surface properties (*i.e.*, hydrophilicity or hydrophobicity, ability to bind specific ligands, etc.). Among the advantages of this type of filtration device is the uniformity and linearity of pores and the high pore to matrix ratio.

In another embodiment, long one-dimensional fibers are incorporated, for example,

30   into paper or cement or plastic during manufacture to provide added wet and dry tensile strength.

In a further embodiment, different nanostructure arrays are impregnated into paper and fabric as anti-counterfeiting markers. In this case, a simple color-linked antibody reaction (such as those commercially available in kits) is used to verify the origin of the material. Alternatively, such nanostructure arrays could bind dyes or other substances,

5   either before or after incorporation to color the paper or fabrics or modify their appearance or properties in other ways.

It will be apparent to one skilled in the art that the nanostructures comprising recombinant gp35 and its derivatives, fragments and analogs include, but are not limited to, other polygonal structures such as octagons, as well as open solids such as tetrahedrons and

10   icosahedrons formed from triangles and boxes formed from squares and rectangles. The range of structures is limited only by the types of angle units and the substituents that can be engineered on the different axes of the rod units. For example, other naturally occurring angles are found in the fibers of bacteriophage T7, which has a 90° angle (Steven et al., *J. Mol. Biol.* 200: 352-365, 1988).

15   The use of bacteriophage tail fiber components in the construction of nanostructures is further described in PCT Publication WO 96/11947, dated April 25, 1996, which is incorporated by reference herein in its entirety.

Additionally, the gp35 proteins, derivatives, fragments and analogs of the invention have use in the study and research of the bacteriophage T4 life cycle.

20

5.9. NANOMETER STRUCTURE FORMATION

Bacteriophage T4 tail fiber proteins gp34, gp35, gp36, and gp37 are produced naturally following infection of *E. coli* cells by intact T4 phage particles. The structure of the T4 bacteriophage tail fiber (illustrated in Figure 1) can be represented schematically as

25   follows (N= amino terminus, C= carboxy terminus): N[34]C - N[gp35]C - N[P36]C - N[P37]C. P34, P36, and P37 homooligomers are stiff and rod-shaped proteins in which two identical β sheets, oriented in the same direction, are fused face-to-face by hydrophobic interactions between the sheets juxtaposed with a 180° rotational axis of symmetry through the long axis of the rod. gp35, by contrast, is a monomeric polypeptide that attaches

30   specifically to the N-terminus of a P36 homooligomer and then to the C-terminus of a P34 homooligomer and forms an angle joint between two rods at an average angle of 137° (±7°) or 156° (±12°).

- 36 -

During T4 infection of *E. coli*, gp37 (the monomeric 109 Kda translation product of *gene 37*) forms the homooligomer P37, with the aid of 2 accessory (chaperon) proteins, gp57 and gp38; this process is believed to initiate near the C-terminus of gp37. Once P37 is formed, the N-terminus of P37 initiates the oligomerization of two gp36 molecules of 23

5   Kda each, in a butt-end joint to form the P36 homooligomer rod. The N-terminus of P36 then attaches to the carboxy terminal region of a gp35 monomer; this interaction stabilizes P36 and forms the flexible angle joint of the tail fiber. The amino terminal region of gp35 then attaches to the C-terminus of P34 (the homooligomerization of which requires the chaperon protein gp57). This regulation of self assembly of the tail fiber by a

10   predetermined, ordered interaction of specific subunits results in the production of a structure of exact specifications from a random mixture of the tail fiber subunit components. Thus, self assembly of the tail fiber is regulated by a predetermined, ordered interaction between specific subunits whereby structural maturation caused by formation of the first subassembly permits interaction with new (previously disallowed) subunits. This results in

15   the production of a structure of exact specifications from a random mixture of the components.

       In one embodiment, the nanostructures of the invention are composed of tail fiber chimers, such as for example, P36-34, which is an oligomer of the fusion protein gp36-34; gp36-34 consists of a portion of gp36 containing the amino terminus fused to a portion of

20   gp34 containing the carboxy terminus. Expression vectors encoding such chimers may be constructed using recombinant technology known in the art. Such chimers have novel functional properties, including but not limited to rod domains and/or N- and C-termini combinations that are different from native tail fiber proteins. Chimers having novel N- and C- termini combinations allow for new patterns for joining different rod segments. For

25   example, polygon nanostructures may be generated using P36-34 chimeric fusion proteins and gp35. The creation of constructs encoding tail fiber fusion chimers, such as P36-34, and their use in generating nanostructures, is further described in PCT Publication WO 96/11947, dated April 25, 1996, which is incorporated by reference herein in its entirety.

       Recombinant expression of the proteins of the present invention in *E. coli* as

30   described above results in the synthesis of large quantities of protein, and allows the simultaneous expression and assembly of different components in the same cells. The methods for scale-up of recombinant protein production are straightforward and widely

known in the art, and many standard protocols can be used to recover native and modified tail fiber proteins from a bacterial culture.

In a preferred embodiment, recombinant gp35 is isolated for use by growing host cells transformed or adsorbed with nucleotide sequence encoding a gp35 protein having the

5   amino acid sequence depicted in Figure 2, operably linked to a heterologous promoter, under conditions in which the gp35 encoding nucleic acid is expressed, and isolating gp35 from the resulting culture by standard methods.

P34, P36-P37, P37 and chimers derived therefrom, such as for example, P36-34, are purified from phage-infected (or recombinant) *E. coli* cultures as mature oligomers. gp35

10   protein, derivatives or analogs thereof are purified as monomers. Standard methods may be utilized to isolate and purify the nanostructure components, these methods include but are not limited to: chromatography on molecular sieve, ion-exchange, and/or hydrophobic matrices; preparative ultracentrifugation; and affinity chromatography, using as the immobilized ligand specific antibodies or other specific binding. For example, if the

15   proteins have been engineered to include heterologous domains that act as ligands or binding sites, the cognate partner may be immobilized on a solid matrix and used in affinity purification. For example, such a heterologous domain can be avidin, which binds to a biotin-coated solid phase.

In an alternative preferred embodiment, several phage tail fiber components, and

20   where necessary, chaperon proteins such as gp57 and gp37 required for homooligomerization, are co-expressed in the same bacterial cells, and sub-assemblies of larger nanostructures are purified subsequent to limited *in vivo* assembly, using the methods enumerated above.

In one embodiment, the purified nanostructure components and/or subassemblies are

25   combined *in vitro* under conditions where assembly of the desired nanostructure occurs at temperatures between about 4°C and about 37°C, and at pH's between about 5 and about 9. For a given nanostructure, optimal conditions for assembly (*i.e.*, type and concentration of salts and metal ions) are easily determined by routine experimentation, such as by changing each variable individually and monitoring formation of the appropriate products.

30   In an alternate embodiment, one or more crude bacterial extracts are prepared, mixed, and assembly reactions are allowed to proceed prior to purification.

In some cases, one or more purified components assemble spontaneously into the desired structure, without the necessity for initiators. In other cases, an initiator is required to nucleate the polymerization of the nanostructure. This offers the advantage of localizing the assembly process (*i.e.*, if the initiator is immobilized or otherwise localized) and of

5   regulating the dimensions of the final structure. For example, rod components that contain a functional P36 homooligomer C-terminus require a functional P37 homooligomer N-terminus to initiate rod formation stoichiometrically; thus, altering the relative amount of initiator and rod component will influence the average length of rod polymer. If the ratio is n, the average rod will be approximately (P37 - P36)n--N-terminus P37-P37 C-terminus.

10   In still other cases, the final nanostructure is composed of two or more components that cannot self-assemble individually, but only in combination with each other. In this situation, alternating cycles of assembly can be staged to produce final products of precisely defined structure.

In one embodiment, polygons are assembled using gp35 and P36-34 chimer.

15   According to this embodiment, gp57 is used to chaperon the homodimerization of gp36-34 to P36-34. P36-34 chimer is added to a solution containing a gp35 initiator that optionally is reversibly immobilized using methods known in the art, so as to allow binding of P36-34 chimer. According to this embodiment, gp35 and P36-34 are administered as a mixture or sequentially to form the desired polygon structure. The type of polygon that is formed

20   using this protocol depends upon the length of rod units and the angle formed by the angle joint. For example, alternating rod units of different sizes can be used. In addition, variant gp35 polypeptides that form angles different than the natural angle can be used, allowing the formation of different regular polygons. Furthermore, for a given polygon with an even number of sides and equal angles, the sides in either half can be of any size provided the

25   two halves are symmetric. The creation of constructs encoding tail fiber fusion chimers, such as P36-34, and their use in generating polygon nanostructures, is further described in PCT Publication WO 96/11947, dated April 25, 1996, which is incorporated by reference herein in its entirety.

When an immobilized initiator is used, it may be desirable to remove the

30   polymerized unit from the matrix after staged assembly. For this purpose specialized initiators are engineered so that the interaction with the first rod component is rendered reversibly thermolabile. For example, where a nanostructure is assembled that is attached

- 39 -

to a solid matrix via gp34, one way in which to detach the nanostructure to bring it into solution is to use a mutant (thermolabile) gp34 that can be made to detach upon exposure to a higher temperature (*e.g.*, 40°C). Such a mutant gp34, termed T4 tsB45, having a mutation at its C-terminal end such that gp34 attaches to the distal tail fiber half at 30°C, but can be

5   separated from it *in vitro* by incubation at 40°C in the presence of 1% SDS (unlike wild-type T4 which are stable under these conditions), has been reported (Seed, 1980, Studies of the Bacteriophage T4 Proximal Half Tail Fiber, Ph.D. Thesis, California Institute of Technology), and can be used. Using a reversibly thermolabile matrix band nanostructure/component, the polymer can be easily separated from the matrix-bound

10 initiator, thereby permitting: easy preparation of stock solutions of uniform parts or subassemblies, and re-use of the matrix-bound initiator for multiple cycles of polymer initiation, growth, and release.

The following examples are intended to illustrate the present invention without limiting its scope.

15

6.   **EXAMPLE: CLONING AND CHARACTERIZATION OF THE BACTERIOPHAGE *gp35* GENE**

As described herein, the present inventor has isolated and characterized the T4 bacteriophage *gene 35*, a gene encoding a tail fiber protein which functions to join the

20 rodlike proximal and distal halves of the bacteriophage tail fibers. Phage T4 *gp35* is located between *gene 34* and *gene 36*. A sequence for *gp35* is available on the NCBI database (NCBI.NIH.GOV) within the sequence T4g34-t (nucleotides 4188-5075; see Figure 3). The NCBI sequence predicts that the *gp35* open reading frame, ORF35 encodes a putative protein having a molecular weight of 32,334 Daltons. However, the present inventor

25 noticed that this deduced molecular weight was discrepant with a reported molecular weight of gp35 as determined by SDS-PAGE of 39,000-40,000 Daltons ("The T4 Book": Molecular Biology of Bacteriophage T4 (1994, Jim Karam editor, ASM Press, Wash. DC, pg. 507 and pg. 514). In addition, the NCBI database predicts a 241 nucleotide open reading frame, ORF34.1, located between *gene 34* and *gene 35* which encodes a protein

30 having a predicted molecular weight of 7,334 Daltons (in a different reading frame from ORF35). The inventor predicted that the NCBI sequence of *gp35* was incorrect and that the two open reading frames, ORFS 34.1 and 35 are actually connected to form a single ORF35

encoding a protein of about 40,000 Daltons. According to this postulation ORF34.1 encodes the N-terminus of gp35.

To prove this hypothesis, the inventor cloned his postulated *gp35* open reading frame by polymerase chain reaction (PCR) of the phage DNA between the 5'-ATG start codon of ORF34.1 and the 3'-TAA stop codon of ORF35, a sequence of approximately 1,120 nucleotides in length, into an inducible expression plasmid, pT7-5, having appropriately situated RNA polymerase and ribosome binding sites and a *lacZ* promoter. Upon induction of expression of the insert from the *lacZ* promoter with IPTG, only one new heavy band (relative to uninduced cells) was apparent on SDS-PAGE, at 41,000 Daltons. There was no visible band at either 32,000 Daltons or 7,000 Daltons.

Sequence analysis of the PCR generated insert revealed that *gp35* contains a single ORF of 1,119 nucleotide pairs having 373 codons, of which 372 encode a protein having a putative molecular weight of 40,096 Daltons. The terminal codon of the *gp35* open reading frame is the ochre stop codon, TAA. This 1,119 nucleotide sequence was compared with the 1,121 nucleotide sequence from the NCBI database using the FASTA program. Six differences were detected between the sequence and that of the NCBI sequence. These six differences are: deletion of the adenine at nucleotide 22 of the NCBI sequence; insertion of a thymine between the adenine at nucleotide 49 and the thymine at nucleotide 50 of the NCBI sequence; deletion of the cytosine at nucleotide 170 of the NCBI sequence; change of nucleotide 238 from a thymine to a cytosine of the NCBI sequence; deletion of the thymine at nucleotide 280 of the NCBI sequence; and change of nucleotide 557 of the NCBI sequence from an adenine to a guanine.

The sequence of the N-terminal 10 residues of the induced protein generated from the expression vector construct were determined to be identical to the first ten residues the inventor predicted for the new *gp35* ORF. The determination of residues 8, 9 and 10 in the induced protein to be phenylalanine, glycine and glutamine, instead of the isoleucine, tryptophan and threonine residues respectively predicted for ORF 34.1 of the NCBI database sequence proves that the new *gp35* ORF sequence is correct and that the adenine located at nucleotide 22 in the NCBI sequence, is the result of a sequencing error and is not actually present in bacteriophage T4 *gene 35*. The inventor has therefore shown that the correct *gp35* sequence is not that previously reported, but actually is a larger protein with a different N-terminus, that is 24% heavier than that predicted from the published sequence.

The correct *gp35* sequence encodes 77 more N-terminal amino acid residues than the NCBI sequence. Additionally, 15 of the first 16 N-terminal residues encoded by the NCBI sequence are incorrect.

Nucleic acid and protein database analysis of the new *gene 35* sequence and its

5   encoded product fails to reveal significant homology with other sequences in the databases.

7.   EXAMPLE: ISOLATION OF THERMOLABILE PROTEINS FOR SELF-ASSEMBLY

A variant (temperature-sensitive) gp35 that permits heat induced separation of the

10   gp35-P36 junction may be formed by mutagenizing the 3' region of *gp35* DNA (encoding the carboxy terminal region of gp35) with randomly doped oligonucleotides. Randomly doped oligonucleotides are prepared during chemical synthesis of oligonucleotides, by adding a trace amount (up to a few percent) of the other three nucleotides at a given position, so that the resulting oligonucleotide mix has a small percentage of incorrect

15   nucleotides at that position (Hutchison et al., 1991, Methods Enzymol. 202:356). The mutagenized DNA fragment is then recombined into T4 phage by infection of the cell containing the mutagenized DNA by a T4 phage containing two amber mutations flanking the mutagenized region. Following a low-multiplicity infection, non-amber phage are selected at low temperature on *E. coli* Su° at 30°C. The progeny of these plaques are

20   resuspended in a buffered solution and challenged by heating at 60°C. At this temperature, wild-type tail fibers remain intact and functional, whereas the thermolabile versions release the P36 units and thus render those phage non-infectious.

At this stage, wild type phage are removed either by adsorbing the wild type phage to sensitive bacteria and sedimenting (or filtering out) the bacteria with the adsorbed wild

25   type phage or by reacting the lysate with anti-gp35-P36 specific antibody, followed by immobilized Protein A and removal of adsorbed wild type phage. Either of these methods leaves the noninfectious mutant phage particles in the supernatant fluid or filtrate, from which they can be recovered. The non-infectious phage lacking terminal gp35-P36 moieties are then urea treated with 6M urea, and mixed with bacterial spheroplasts to permit

30   infection at low multiplicity whereupon they replicate at low temperature and release progeny. Alternatively, infectious phage are reconstituted by *in vitro* incubation of the mutant phage with wild type P36 at 30°C; this is followed by infection of intact bacterial

cells using a standard protocol. The latter method of infection specifically selects mutant phage in which the thermolability of the gp35-P36 junction is reversible.

Using either method of infection, the phage populations are subjected to multiple rounds of selection, after which individual phage particles are isolated by plaque

5 purification at 30°C. Finally, the putative mutants are evaluated individually for: loss of infectivity after incubation at high temperatures (40-60°C), as measured by a decrease in titer; loss of P36 after incubation at high temperature, as measured by decrease in binding of gp35-P36-specific antibody to phage particles; and morphological changes in the tail fibers after incubation at high temperatures, as assessed by electron microscopy.

10 After the mutants are isolated and their phenotypes confirmed, the *gp35* gene is preferably sequenced. If the mutations localize to particular regions or residues, those sequences are preferably targeted for site-directed mutagenesis to optimize the desired characteristics.

Subsequently, the mutant *gene 35* is cloned into an expression plasmid and

15 expressed individually in *E. coli*. The mutant gp35 protein is then purified from bacterial extracts and used in *vitro* assembly reactions.

In a similar fashion, gp35 variants can be isolated that exhibit a thermolabile interaction with P34. In contrast to the localized mutagenesis described above, the screen for gp35 mutants exhibiting a thermolabile interaction with P34 involves random doped

20 oligonucleotide mutagenesis of the entire *gp35* gene. Mutants generated according to the experimental protocol described above are incubated at a high temperature, resulting in the loss of the entire distal half of the tail fiber (*i.e.*, gp35-P36-P37) in the thermolabile mutants. Wild-type phage (and distal half-fibers from thermolabile mutants) are then separated from thermolabile mutant phage that have been inactivated at high temperature (but still have

25 proximal half tail fibers attached) by precipitating both the distal half-fibers and the phage particles containing intact tail fibers with any of the anti-distal half tail-fiber antibodies and protein-A beads. Mutant phage remaining in the supernatant are then reactivated by incubation at low temperature with bacterial extracts containing wild type intact distal half fibers. The thermolabile gene 35 mutants grown at 30°C can be tested for reversible

30 thermolability by inactivation at 60°C and reincubation at 30°C. Inactivation is performed on a concentrated suspension of phage, and reincubation at 30°C is performed either before

or after dilution. If phage are successfully reactivated before, but not after dilution, this indicates that their gp35 is reversibly thermolabile.

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those
5 described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.
10

15

20

25

30